

A Faster and Cheaper Method to Detect Agents of Disease

Livermore's pathogen detection technology identifies nearly 6,000 different microbes within 24 hours.

SPEEDY, accurate identification of pathogens—the viruses, bacteria, and fungi that cause disease—is becoming increasingly important. Infectious diseases pose a growing threat to public health due to population growth, international air travel, bacterial antibiotic resistance, and other factors. In addition, forensic experts are increasingly concerned that pathogens, perhaps genetically engineered, could be released deliberately by terrorist organizations or rogue states.

While several techniques exist for identifying pathogens via their genetic code, most of these methods are too costly or slow to efficiently analyze clinical and environmental samples that may contain hundreds or even thousands of different microbes. Lawrence Livermore

researchers have developed a technology that rapidly identifies any known microbe whose genetic code has been sequenced. Called the Lawrence Livermore Microbial Detection Array (LLMDA), the technology combines innovative bioinformatics (the discipline of analyzing biological data using computational tools) with a tiny device called a microarray.

Livermore scientists analyzed the genetic code of every microbe that has been sequenced (about 6,000 species and strains in all) and then selected the roughly 360,000 most important genetic markers. In one microarray configuration, 360,000 probes—short stretches of DNA or RNA that complement the isolated genetic markers—are arrayed in a microscopic square grid on a

2.5- by 7.5-centimeter glass slide. When a fluorescently labeled fluid sample containing the genetic material of microbes contacts the microarray's probes, only the squares with DNA or RNA unique to a particular organism are activated. The activated squares produce a fluorescent pattern, from which species present in the sample are identified. In this way, multiple pathogens are detected simultaneously, with typical processing times of less than 24 hours. The current-generation LLMDA can identify 3,111 viruses, 1,967 bacteria, 94 protozoa, 136 fungi, and 126 archaea (primitive bacteria).

In trials for government agencies, international researchers, and health-product companies, LLMDA has accurately



and rapidly identified bacterial and viral pathogens present in human and animal clinical samples, environmental samples, and product samples. Government agencies and private research centers are collaborating with Livermore's LLMDA team to identify viruses and bacteria that are correlated with high cancer risk, vaccine safety, and defense against a bioterrorist attack. If widely adopted, LLMDA could allow professionals in medicine, pharmaceuticals, law enforcement, product and food safety, public health, animal health, the military, and global disease surveillance to detect within 24 hours any virus or bacteria that has been sequenced and included among the array's probes.

Current sponsors of the pathogen detection effort include the Department of

Defense and the Department of Homeland Security (DHS). Collaborators include the University of California at San Francisco; Blood Systems Research Institute in San Francisco, California; Moffitt Cancer Center in Tampa, Florida; University of Texas Medical Branch at Galveston; National Institute for Public Health and the Environment in Bilthoven, Netherlands; University of California at Davis; U.S. Food and Drug Administration (FDA); Centers for Disease Control and Prevention (CDC); Naval Medical Research Center; and Marine Mammal Center in Sausalito, California.

LLMDA was licensed in 2012 to MOgene, LC, a U.S.-based supplier of DNA microarrays and instruments. The Statens Serum Institut in Denmark has also licensed

the device for use as its primary virus-screening tool. A number of other industrial collaborators have expressed interest in licensing LLMDA or in having samples analyzed with the device. Licenses include analysis software. However, licensees must provide their own computer for analyzing results.

Computer scientist Thomas Slezak, who leads Livermore's pathogen bioinformatics team, conceived of LLMDA in 2003 when he learned of an important advance in directly synthesized DNA microarray technology. While microarrays have been around for several years, their use has been limited because each probe was restricted to about 25 DNA or RNA bases. A technology breakthrough has permitted much larger probes (60 or more DNA or

RNA bases). About 10-percent variation in the probes has proven to be acceptable. If the DNA in a sample differs from a probe by up to six bases, the probe can still detect a match. “The earlier, smaller probes were too sensitive to withstand even a single mismatch variation,” says Slezak. “Because pathogen strains circulating in nature contain variations in DNA that can differ from the strains already sequenced, it was imperative for us to use a technology that was robust to natural—or engineered—variation.”

“Early sponsors were difficult to find,” he adds, “because the concept of simultaneously testing for thousands of microbes seemed unrealistic.” As a result, work on the first-generation LLMDA began in 2007 as a Laboratory Directed Research and Development project.

The current development team includes biologist Crystal Jaing, who leads the microarray laboratory work and manages collaborations; bioinformaticist Shea Gardner, who designed the array and probe selection effort; bioinformaticist Kevin McLoughlin, who designed the analysis

software; and biologists James Thissen and Nicholas Be, who perform LLMDA experiments. Jaing, Thissen, and Be are part of the Applied Genomics Group of the Biosciences and Biotechnology Division, which has extensive experience in biodetection. Gardner and McLoughlin belong to Slezak’s 12-member pathogen bioinformatics team, which develops advanced assays for detecting pathogens. In 2000, this team created the world’s first automated pathogen DNA-signature detection system.

How LLMDA Works

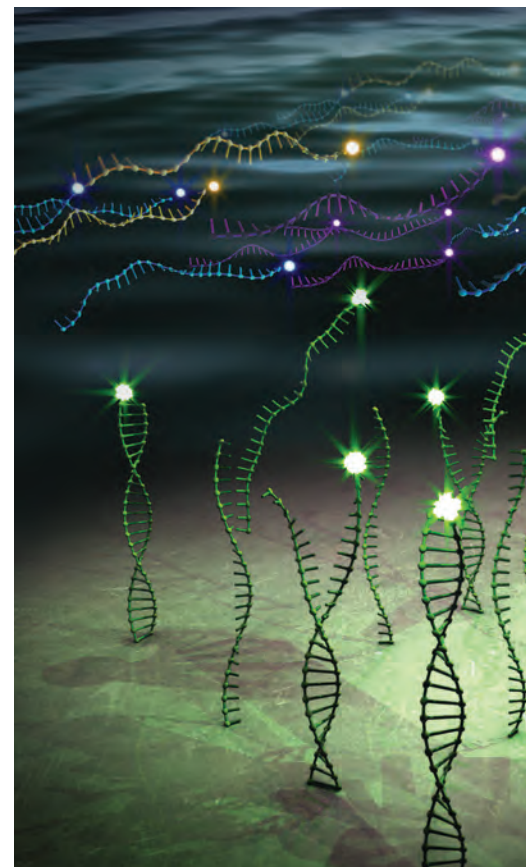
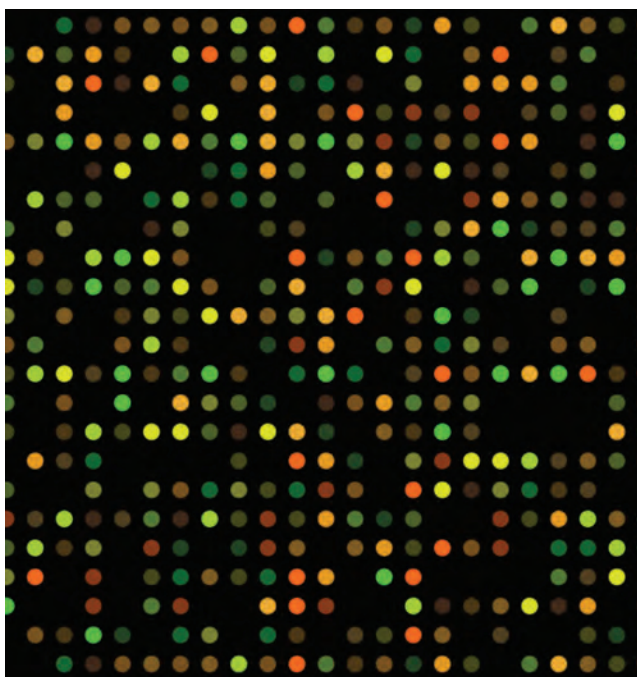
In the past few years, microarrays have attracted increased interest from clinicians, government agencies, and disease researchers because of their ability to analyze foods, pharmaceuticals, and complex clinical and environmental samples. While other nucleic-acid-based microarrays can detect certain classes of microbes, such as viruses, LLMDA is the only one that provides simultaneous characterization of both bacteria and viruses. Says Jaing, “LLMDA is sensitive

and specific. It can detect very low concentrations of a particular microbe.”

The key principle behind LLMDA is hybridization between two complementary sequences of nucleic acids. The probes have sequences corresponding to segments of an organism’s genome and are sprayed onto a glass slide in a manner similar to ink-jet printing. The probes can also be built using chemical photodeprotection technology. Each glass slide contains one or more aggregates of probes, and each aggregate features hundreds of thousands of squares arranged in a grid. Several dozen squares on the grid contain probes that correspond to unique genetic sequences from a single organism.

Several glass slide configurations are possible, ranging from one square grid containing all 360,000 probes for detecting any microbe previously sequenced, to

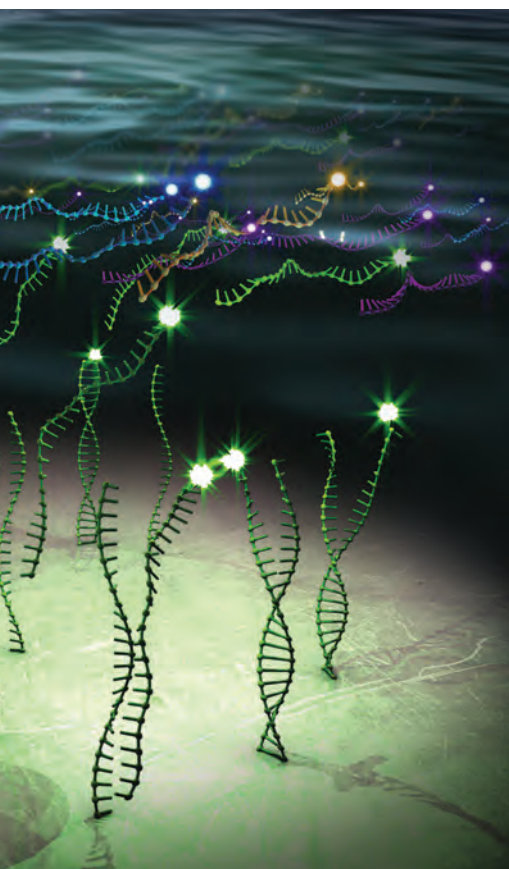
When a fluorescently labeled sample of fluid containing the genetic material of microbes contacts the Lawrence Livermore Microbial Detection Array’s (LLMDA’s) probes, only the squares with DNA or RNA unique to a particular organism are activated. The activated squares produce a fluorescent pattern, from which species present in the sample are identified. In this way, multiple pathogens are detected simultaneously, with typical processing times of less than 24 hours.



12 square grids on one slide, each with 135,000 probes. The latter is designed for human clinical purposes and allows for specimens from a dozen patients to be analyzed simultaneously.

The detection process begins with DNA and RNA (from microbes) extracted from a clinical or environmental sample. The DNA and RNA are amplified, if needed (for example, if a bacterium concentration is expected to be quite low, as from an aerosol sample). The genetic material is fragmented and labeled with a fluorescent dye and then applied to the microarray at 42°C for several hours, allowing the fragments to hybridize to their

In the rendering below, viral RNA fragments fluorescently tagged from a sample hybridize to LLMDA probes. (Rendering by Kwei-Yu Chu.)



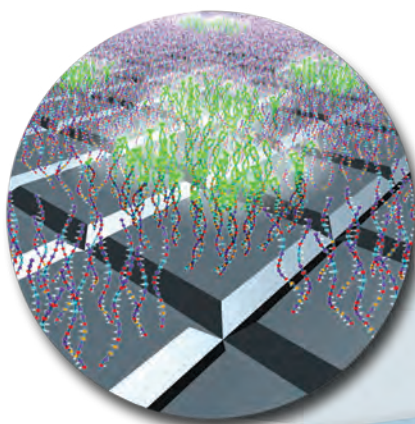
complementary probes. After unbound genetic material is washed off, only strongly hybridized pairs remain, which fluoresce brightly. An automated system, guided by Livermore software, examines the pattern of squares that light up to identify the virus or bacterium, sometimes down to the strain level.

Jaing notes that when compared with the two main microbe detection technologies—polymerase chain reaction (PCR) and DNA sequencing—LLMDA is mid-range in cost, processing time, and sensitivity. PCR analysis is relatively inexpensive, fast, and sensitive for known organisms, but it can detect no more than about 50 different organisms at one time. The PCR assays are too limited for analyzing the thousands of species of pathogens that have been sequenced. In contrast, LLMDA can identify previously sequenced bacteria and viruses as well as new pathogens containing DNA sequences similar to those previously identified in other pathogens. At the other end, DNA sequencing provides the most comprehensive information about

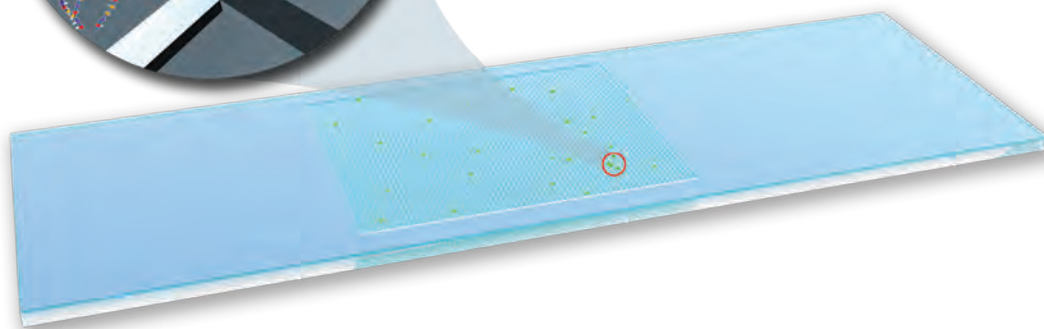
pathogens but is costly and can take several days to complete. LLMDA is much faster and cheaper than sequencing.

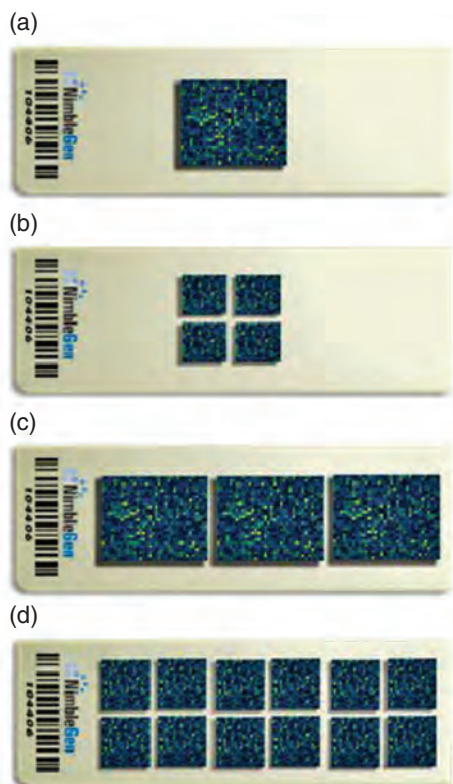
Designing 360,000 Probes

Key to the Livermore technology is the specificity of its 360,000 probes, each selected to help detect one microbe or a set of microbes. A probe consists of typically 50 to 65 nucleotides based on a region of RNA or DNA from the available viral, bacterial, fungal, protozoan, and archaeal genomes. More than 100 conserved and unique probes on average are selected per species. Says Gardner, “Unique parts discriminate one species or strain from another. Conserved parts are the same in all strains. Because conserved parts are so important to the organism, they don’t mutate away.” Gardner routinely updates the probes with new sequences of bacteria, viruses, and other microorganisms published in public databases as well as new sequences obtained from collaborators. However, the process is not yet automated and can take weeks to perform on an entire year’s collection of new data.



LLMDA features aggregates of probes arranged in a square grid. The grid has hundreds of thousands of squares, and each square holds millions of copies of a single probe. Several dozen squares on the grid contain probes that correspond to unique genetic sequences from a single organism. (Rendering by Sabrina Fletcher.)





Several microarray configurations are possible on a single slide, including (a) a single square array containing all 360,000 probes, (b) four arrays of 72,000 probes each, (c) three arrays of 720,000 probes, and (d) for efficient human clinical use, 12 arrays on one slide, each with 135,000 probes.

Gardner developed a “software pipeline” to analyze every sequenced microbe and extract stretches of DNA and RNA that might make good probes. The task required hundreds of thousands of central-processing-unit hours (over about 45 days) using powerful Livermore cluster computers. The pipeline requires more than a dozen steps, each involving a different algorithm. For example, algorithms are used to calculate unique regions of nucleotides (while removing nonunique regions), search out conserved regions within a family, and predict how “sticky” a probe will be to its nucleic-acid complement in the sample.

The algorithms seek to balance the goals of conservation and uniqueness, prioritizing sequences that were conserved within the family of the targeted organism and are unique relative to other families. For example, selecting probes that correspond only to what makes this year’s influenza virus unique will fail to identify the virus when it mutates, as viruses tend to do. As a result, Gardner looks for some stretches of genetic material that are conserved. The use of multiple probes also makes it possible to discriminate between strains of the same species. Probes are designed to have no significant matches to the human genome sequence.

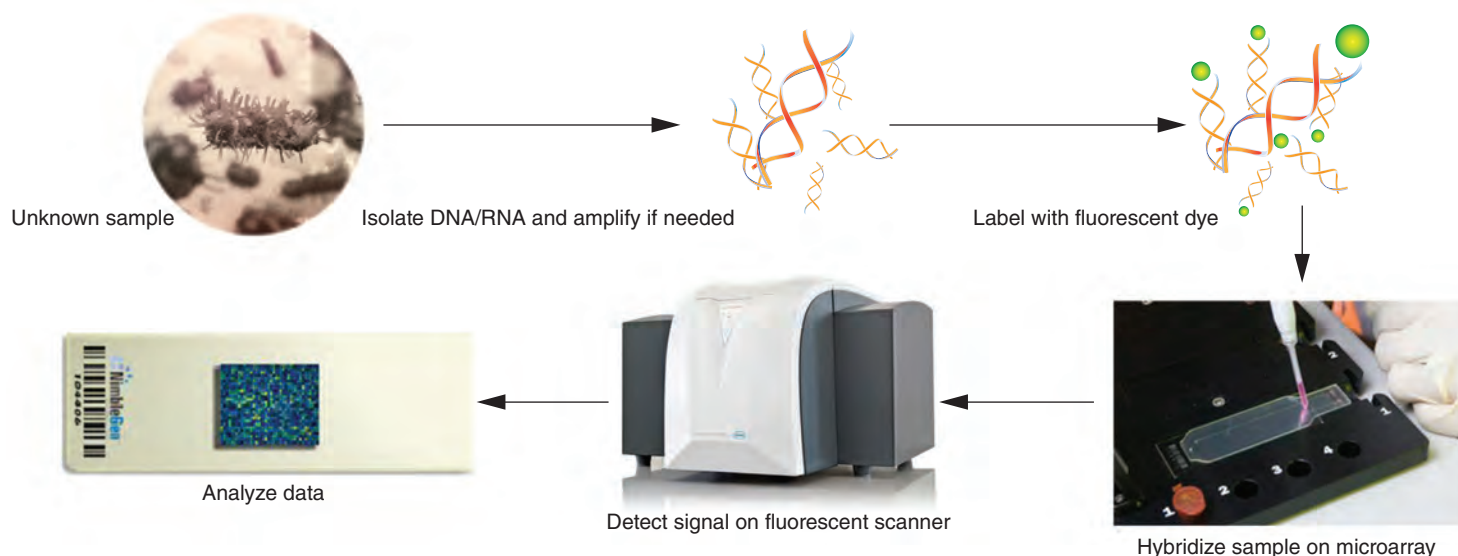
In addition, a set of 2,600 negative control probes has sequences that are randomly generated, but with length and content of cytosine and guanine (two nucleic-acid bases) that match those of the target-specific probes. The negative control probes establish a built-in background rate of fluorescence for each microarray analysis.

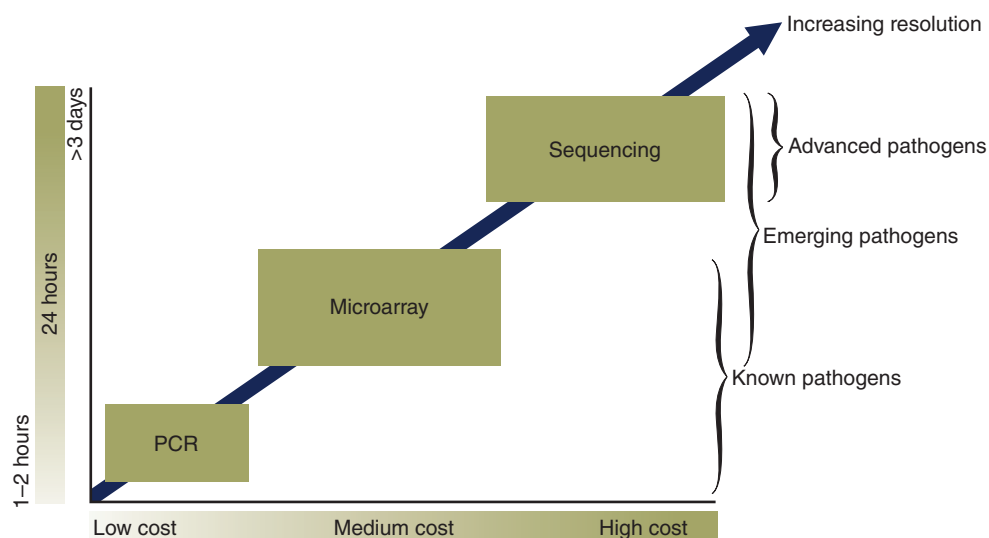
In all, the algorithms scan more than 20 billion bases comprising the genetic code of close to 6,000 species. Looked at another way, the task corresponds to searching through more than 666 million books and locating within each book at least 60 phrases unique to that book. Together, the probes contain data that make up a 60-megabyte file.

Detection Analysis

Analyzing the microarray results requires a workstation with 200 gigabytes

In the LLMDA analysis process, DNA and RNA are extracted from a sample, labeled with a fluorescent dye, and hybridized with the probes arranged on a microarray. After unbound genetic material is washed off, only strongly hybridized pairs remain, which fluoresce brightly. The microarray is then scanned, and the data are analyzed.





When compared with polymerase chain reaction (PCR) and DNA sequencing, microarrays such as LLMDA are mid-range in cost, processing time, and sensitivity. PCR analysis is relatively inexpensive, fast, and sensitive for known organisms, but it can detect no more than about 50 different organisms at one time. In contrast, LLMDA can identify thousands of previously sequenced bacteria or viruses, including new pathogens containing DNA sequences already identified in other pathogens. Although DNA sequencing provides the most comprehensive information about pathogens, it is costly and takes much longer to complete than LLMDA.

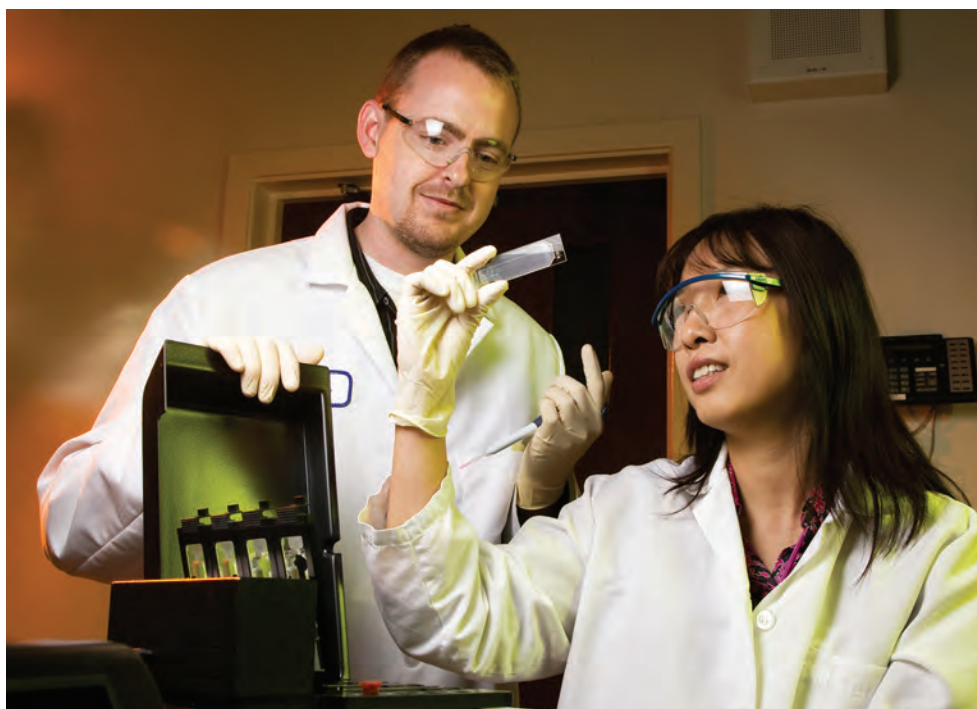
of memory and 12 processors as well as the Livermore-developed analysis algorithm that makes sense of the voluminous data produced by the scan. The analysis begins when the slide containing the hybridized probes is placed in a scanner. A laser scans across the slide's surface, and a photodetector picks up the fluorescing squares. Within a few minutes, an enormous image file of 100 megabytes is built up. Commercial software analyzes the image and produces a file quantifying the fluorescent intensity at each spot.

To identify the organisms that best explain the probe intensities recorded on the image file, McLoughlin developed the composite likelihood maximization algorithm. The computer program searches repeatedly through the database of all sequenced microbial genomes, at each iteration choosing the most likely pathogen to match the fluorescent pattern. In the first iteration, it looks for the target genome that explains the largest portion of the

detected probe signals. In each subsequent iteration, the algorithm chooses the organism that explains the largest part of the signal not already explained by the first target. Computer scientists call this kind of algorithm "greedy" because it is always grabbing for the best explanation, then the next most likely explanation, and so on.

The microarray results, ready in 10 to 20 minutes, display a list of predicted targets organized by viral or bacterial family that reflects the most probable organisms corresponding to the detailed fluorescent pattern. For Livermore users, the results are automatically available online via a Web-based interface. After logging on, users can query any LLMDA test and request analysis in various ways. For example, a user may request a list of only the viruses present in a particular sample. When the analysis is complete, the user receives an e-mail with a link to the results.

A system for analyzing LLMDA results has been deployed at DHS's microbial



Biologists James Thissen (left) and Crystal Jaing work with an LLMDA slide that is used to detect nearly 6,000 different microbes.

forensics center. Another is deployed at the Statens Serum Institut, where the system is reportedly identifying viruses more efficiently than numerous PCR tests. This year, the Livermore team is supporting a large-scale evaluation of LLMDA at both CDC and the U.S. Army Medical Research Institute of Infectious Diseases.

Health, Biodefense Applications

In the area of biodefense, the LLMDA platform provides a “safety net” for rapid detection of known pathogens that might not be watched for in the first line of defensive systems (for example, BioWatch or the CDC’s Laboratory Response Network). It also offers an orthogonal confirmation to PCR recognition of a

known pathogen. And, unlike PCR, the platform can provide a rough guess as to the closest strain.

The Livermore team has developed other customized biodefense arrays for DHS that can recognize genes involved in known virulence or antibiotic resistance pathways, detect genetic engineering vectors, and provide very high-resolution phylogenetic strain typing for many key bacterial and viral threat agents. These arrays have been transitioned to the DHS National Biodefense Analysis and Countermeasures Center.

Jaing says LLMDA could prove particularly useful to CDC for tracking emerging diseases, whether they are recently discovered or previously known but causing a new outbreak, such as severe acute

respiratory syndrome, or SARS. CDC is now testing a hierarchy for diagnosing unknown pathogens. The center will first use PCR to identify a microbe. If that technique is unsuccessful, the center will turn to LLMDA. If LLMDA does not identify it, meaning the microbe’s genetics have not yet been sequenced, the center will use DNA sequencing. CDC has sent researchers to Livermore for training with LLMDA.

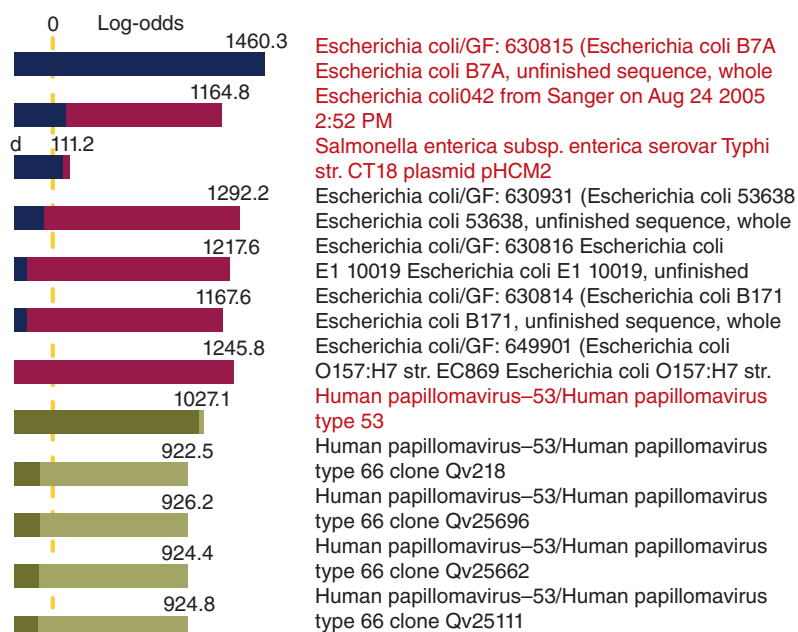
The Livermore team is also working with Department of Defense agencies such as the Naval Medical Research Center to identify pathogens in combat-wound samples and the microbial pattern predictive of wound infection and healing. LLMDA has detected clinically relevant pathogens from wound samples more rapidly and accurately than traditional microbiological techniques such as culturing a sample to see if a bacterial colony forms.

In addition, LLMDA has identified viruses and bacteria that are correlated with a high cancer risk to aid in early detection and prevention strategies. For example, LLMDA detected human papillomavirus 16 from cancer samples. The virus causes about 70 percent of cervical cancers and is the leading cause of oral cancer.

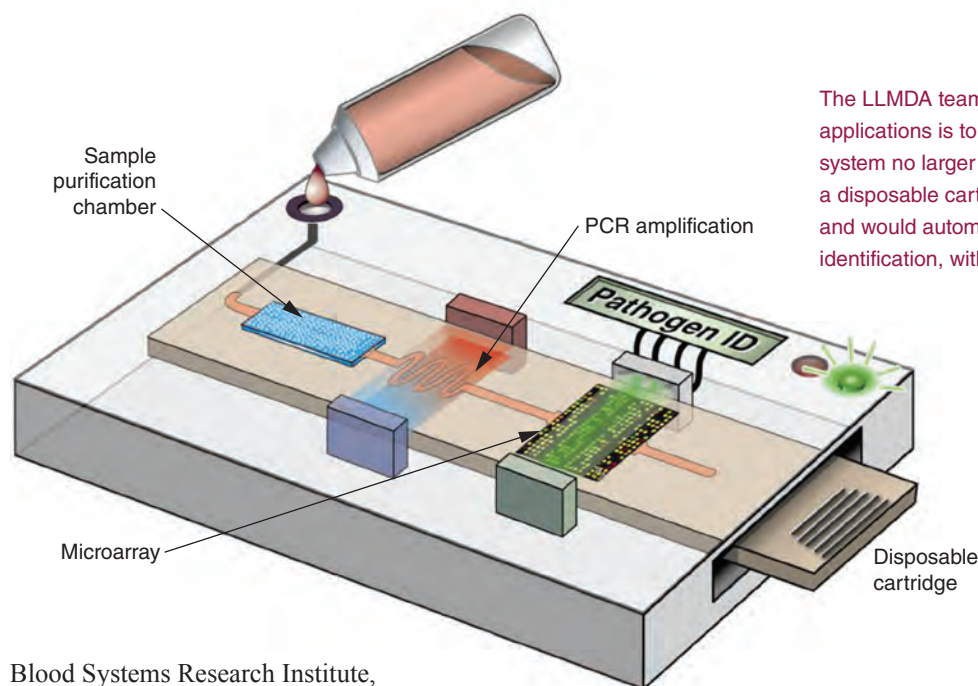
Finally, samples of DNA associated with the remains of Black Death victims from the Middle Ages are being studied with LLMDA to identify the pathogens that may have caused the disease. The LLMDA research team, along with Livermore bioscientist Monica Borucki, is conducting the study in collaboration with colleagues at McMaster University in Canada.

Ensuring Vaccine Safety

An important potential application for LLMDA is ensuring vaccine safety by testing live, attenuated viral vaccines for any potential contaminant viruses. The process of creating vaccines uses components derived from animals and runs the risk of contaminating the vaccine with other viruses (called adventitious). Working with the San Francisco-based



The composite likelihood maximization algorithm identifies the organisms that best explain the probe intensities recorded on an image file. The analytic results, ready in 10 to 20 minutes, are listed by viral and bacterial family in order of the most probable organisms that correspond to the detailed fluorescent pattern. Pathogens are listed within families in decreasing order of likelihood (log-odds) scores. Targets predicted most likely to be present are indicated in red text. The lighter- and darker-colored portions of the bars represent the unconditional and conditional scores, respectively. That is, the darker-colored portion shows the contribution from a target that cannot be explained by another, more likely target above it, while the lighter-colored portion illustrates that some very similar targets share a number of probes, so multiple targets may be consistent with the hybridization signals.



The LLMDA team's goal for biodefense and human and health applications is to develop an extremely compact, fully integrated system no larger than a cell phone. This portable device would feature a disposable cartridge holding an environmental or clinical sample and would automate every task, from sample preparation to pathogen identification, within 1 hour.

Blood Systems Research Institute, Livermore researchers used LLMDA in 2011 to evaluate seven live, attenuated viral vaccines: oral poliovirus, rubella, measles, yellow fever, varicella-zoster (herpes), multivalent measles/mumps/rubella, and rotavirus. The institute's team tested for the presence of adventitious viruses by sequencing all genetic material in the vaccines. LLMDA confirmed the institute's findings (without knowing the sequencing results), but much faster and at much lower cost.

Obstacles Still Remain

Many experts foresee that microarrays will eventually become a popular means for identifying pathogens present in clinical samples as well as ensuring quality control for food and biological products. An LLMDA system containing probes for all human pathogens could replace hundreds of individual PCR assays and eliminate the need for a clinical hypothesis regarding a suspected pathogen. Jaing notes, however, that the capabilities of LLMDA are limited by the genome sequence information available. Many species and strains of known microbial pathogens have not yet been sequenced.

Slezak adds that nontechnical obstacles exist to widespread adoption of LLMDA such as issues associated with medical product regulation, intellectual property, the culture of U.S. medicine, and health insurance. FDA does not have defined protocols for evaluating a device that simultaneously tests for thousands of microbes. Likewise, the U.S. Patent Office does not have experience dealing with such a device.

Also, U.S. physicians do not currently focus on the precise diagnosis of an infectious disease. "The task of the primary physician is to determine if the infection is viral or bacterial. If the infection is bacterial, antibiotics are prescribed," says Sleazak. He also notes, "The basic insurance model is one patient, one test, one result, one payment. Insurance companies are unsure how to deal with a test that can detect many different pathogens." However, further enhancements to LLMDA, combined with organizational innovations in medical care, could lead to the widespread use of the device and improve diagnosis speed while reducing costs.

The Livermore team's long-term goal for biodefense and human and animal health applications is to develop an extremely compact, fully integrated system no larger than a cell phone. This portable device would feature a disposable cartridge holding an environmental or clinical sample and would automate every task, from sample preparation to pathogen identification. Such a device would be extremely useful if, for example, a white powder were discovered that looked suspiciously like anthrax. First responders could expect highly reliable results within 1 hour. (Current rapid field tests for pathogens have unacceptably high error rates.)

LLMDA encompasses clever algorithms, microtechnology, and innovative thinking. It is poised to provide a powerful new weapon to fight disease, ensure the safety of vaccines and food products, and provide increased protection from a bioattack.

—Arnie Heller

Key Words: bacteria, bioattack, bioinformatics, Blood Systems Research Institute, Centers for Disease Control and Prevention (CDC), composite likelihood maximization algorithm, Department of Defense, Department of Homeland Security (DHS), Lawrence Livermore Microbial Detection Array (LLMDA), microarray, pathogen, polymerase chain reaction (PCR), sequencing, Statens Serum Institut, U.S. Food and Drug Administration (FDA), virus.

For further information contact Crystal Jaing (925) 424-6574 (jaing2@llnl.gov).